Preprint ANL/MCS-P1742-0410

Achieving Higher Frequencies in Large-Scale Nonlinear Model Predictive Control

Victor M. Zavala and Mihai Anitescu

***Abstract*— We present new insights into how to achieve higher frequencies in large-scale nonlinear predictive control using *truncated-like* schemes. The basic idea is that, instead of solving the full nonlinear programming (NLP) problem at each sampling time, we solve a single, truncated quadratic programming (QP) problem. We present conditions guaranteeing stability of the approximation error derived through this type of scheme using generalized equation concepts. In addition, we propose a preliminary scheme using an augmented Lagrangian reformulation of the NLP and projected successive over relaxation to solve the underlying QP. This strategy enables early termination of the QP solution because it can perform linear algebra and active-set identification tasks simultaneously. A simple numerical case study is used to illustrate the developments.**

I. PROBLEM STATEMENT

Consider a nonlinear predictive control (NMPC) problem of the form

$$\min_{u(\tau)} \int_t^{t+T} \psi\left(z(\tau), u(\tau)\right) d\tau \tag{1a}$$

$$\text{s.t. } \dot{z}(\tau) = \phi(z(\tau), u(\tau)),\ \tau \in [t, t+T] \tag{1b}$$

$$z(\tau) \geq 0,\ u(\tau) \geq 0,\ \tau \in [t, t+T] \tag{1c}$$

$$z(t) = \bar{z}(t), \tag{1d}$$

where t is the time dimension, T is the prediction horizon, $u(\cdot) \in \Re^{n_u}$ are the control trajectories, and $z(\cdot) \in \Re^{n_x}$ are the *model* state trajectories. The nonlinear mapping $\phi(\cdot,\cdot) : \Re^{n_z \times n_u} \to \Re^{n_z}$ is the model, and $\psi(\cdot,\cdot) : \Re^{n_z \times n_u} \to \Re$ is the cost. This problem is parametric in the *system* states $\bar{z}(\cdot)$ which we assume evolve according to the dynamics

$$\dot{\bar{z}}(t) = \phi(\bar{z}(t), u(t)) + w(t), \tag{2}$$

where $w(\cdot)$ is a disturbance process reflecting unmodelled dynamics or noise. In an ideal scenario, we seek to solve the NMPC problem in *real-time* in order to reject the disturbances. A fundamental challenge is that, if the NMPC problem is not solved at a frequency consistent with the system dynamics, the disturbances will accumulate over time and compromise the system stability. In many applications achieving the desired frequencies is not possible due to the computational complexity of the NMPC problem.

Currently, the NMPC problem is typically solved by casting this as a nonlinear programming (NLP) problem. Here, we consider a parametric NLP of the form

$$\min \quad f(x,t), \quad \text{s.t.} \quad c(x,t) = 0, \quad x \geq 0, \tag{3}$$

V. M. Zavala and M. Anitescu are with the Mathematics and Computer Science Division at Argonne National Laboratory, Argonne, IL 60439 {vzavala,anitescu}@mcs.anl.gov

where $x \in \Re^n$ are the decision variables that include discretized controls and states and, with some abuse of notation, $t \in \Re$ will be used to represent the time-evolving data (e.g., system states). A solution of this problem satisfies the parametric Karush-Kuhn-Tucker (KKT) system:

$$x^T \nabla_x \mathcal{L}(w,t) \geq 0,\ c(x,t) = 0,\ x \geq 0. \tag{4}$$

This is a nonlinear complementarity system. The Lagrange function is defined as

$$\mathcal{L}(w,t) = f(x,t) + \lambda^T c(x,t), \tag{5}$$

where $\lambda \in \Re^m$ are Lagrange multipliers and $w^T = [x^T, \lambda^T]$. We denote the solution of this problem as $w^*(t)$. Under certain regularity conditions, this solution forms a continuous but *non-smooth* manifold [6]. Non-smoothness arises as a result of variables hitting and moving away from the bounds at subsequent times (active-set changes).

In an effort to achieve higher frequencies, several studies [3], [8], [10] have proposed to solve a single quadratic programming (QP) problem per sampling time. The QP derived from (3) has the form

$$\min_{\Delta x \geq -\bar{x}_{t_k}} \nabla_x f(\bar{x}_{t_k}, t_{k+1})^T \Delta x + \frac{1}{2} \Delta x^T H(\bar{w}_{t_k}, t_k) \Delta x \tag{6a}$$

$$\text{s.t.} \qquad c(\bar{x}_{t_k}, t_{k+1}) + J(\bar{x}_{t_k}, t_k) \Delta x = 0, \tag{6b}$$

where $H = \nabla_{xx} \mathcal{L}$ is the Hessian of the Lagrange function and $J = \nabla_x c$ is the Jacobian of the equality constraints. The basic idea is to linearize the KKT conditions at the current point $\bar{w}_{t_k}$, perturb the data $t_{k+1} = t_k + \Delta t$, solve the QP to obtain Δw_{t_k}, and obtain $w_{t_{k+1}} = \bar{w}_{t_k} + \Delta_w$. The hope is that $\bar{w}_{t_k}$ provides a good approximation to $w^*(t_k)$ and that it converges as $k \to \infty$. An interesting observation arising in NMPC is that the sampling time can be reduced as we decrease the solution time. In the limit, we can expect that $\|\bar{w}_t - w_t^*\| \to 0,\ \forall t$ as $\Delta t \to 0$. The traditional approach of converging the NLP to a high degree of accuracy considerably limits the attainable Δt. Therefore, it seems preferable to obtain a fast and sufficiently accurate solution in order to reject disturbances and to *keep neighboring problems close to each other*. Notice, however, that the minimum achievable sampling time Δt is restricted by the solution time of the QP itself. Consequently, some questions arise: For finite δt, is it possible to guarantee that the error $\|\bar{w}_{t_k} - w_{t_k}^*\|$ remains stable? Can we guarantee $\|\bar{w}_{t_k} - w_{t_k}^*\| \to 0$ as $k \to \infty$? How can we accelerate the QP solution to reduce Δt? In this work, we address stability by

using generalized equation concepts (Section II). In addition, we propose a scheme to terminate the QP solution early and still guarantee stability of the error (Section IV). We illustrate these developments using a simple numerical case study (Section V).

II. GENERALIZED EQUATIONS

The key observation driving our analysis is the fact that the complementarity system (4) can be posed as a parametric *generalized equation* (GE) of the following form: For a given $t \in T \subseteq \Re$, find $w \in W \subseteq \Re^n$ such that

$$0 \in F(w,t) + \mathcal{N}_K(w). \tag{7}$$

Here, $F : W \times T \to Z$ is a continuously differentiable mapping in both arguments with

$$F(w,t) = \begin{bmatrix} \nabla_x \mathcal{L}(w,t) \\ c(x,t) \end{bmatrix}, \tag{8}$$

and $K = \Re^n_+ \times \Re^m \subseteq W$ is a polyhedral convex set and $\Re^n_+$ is the non-negativity orthant. We define the derivative mapping $F_w(w,t) := \nabla_w F(w,t)$ and assume that it is Lipschitz in both arguments with constant $L_{F_w}, \forall w \in W, t \in T$. The multifunction $\mathcal{N}_K : W \to 2^Z$ is the normal cone operator

$$\mathcal{N}_K(w) = \begin{cases} \{\nu \in W \mid (w-y)^T \nu \geq 0,\ \forall\, y \in K\} & \text{if } w \in K \\ \emptyset & \text{if } w \notin K. \end{cases}$$

Our goal is to create a discrete-time scheme $\bar{w}_{t_k}$ providing a fast but stable approximation of the solution of (7) $w^*_{t_k}$. To achieve this, we will perform a single truncated Newton iteration for the generalized equation per time step.

A. The Nonlinear Equation Case

A good intuition as to why a truncated scheme is sufficient to track the solution manifold can be easily explained by considering the case without inequality constraints or, equivalently, when $K = \Re^n$ and $F(w,t) = 0$. In this case, standard calculus results can be used to establish error bounds. This approach has been followed in [3], [10]. In this section, we perform an informal analysis in order to motivate the results of later sections. In the absence of inequality constraints, the approximate scheme $\bar{w}_{t_k}$, $k > 0$, can be obtained from the recursive solution of the truncated linear Newton system:

$$r_\epsilon = F(\bar{w}_{t_k}, t_{k+1}) + F_w(\bar{w}_{t_k}, t_k)(\bar{w}_{t_{k+1}} - \bar{w}_{t_k}) \tag{9}$$

where r_ϵ is the solution residual satisfying $\|r_\epsilon\| \leq \kappa_\epsilon > 0$. We assume by now that the linearization point $\bar{w}_{t_k}$ satisfies $\|\bar{w}_{t_k} - w^*_{t_k}\| \leq \kappa_r$ where $F(w^*_{t_k}, t_k) = 0$. In addition, we assume that the solution manifold is Lipschitz continuous (see Theorem 1) such that $\|w^*_{t_{k+1}} - w^*_{t_k}\| \leq L_w \Delta t$ with $\Delta t = t_{k+1} - t_k$ and $\kappa_r, L_w > 0$. We need to establish conditions leading to stability of the approximation error in the sense that

$$\|\bar{w}_{t_k} - w^*_{t_k}\| \leq \kappa_r \quad \Rightarrow \quad \|\bar{w}_{t_{k+1}} - w^*_{t_{k+1}}\| \leq \kappa_r.$$

From the mean value theorem we have that

$$\begin{aligned} 0 &= F(w^*_{t_{k+1}}, t_{k+1}) \\ &= F(w^*_{t_k}, t_{k+1}) \\ &\quad + \int_0^1 F_w\left(w^*_{t_k} + \chi(w^*_{t_{k+1}} - w^*_{t_k}), t_{k+1}\right)(w^*_{t_{k+1}} - w^*_{t_k}) d\chi \end{aligned} \tag{10}$$

and

$$\begin{aligned} F(w^*_{t_k}, t_{k+1}) &= F(\bar{w}_{t_k}, t_{k+1}) \\ &\quad + \int_0^1 F_w\left(\bar{w}_{t_k} + \chi(w^*_{t_k} - \bar{w}_{t_k}), t_{k+1}\right)(w^*_{t_k} - \bar{w}_{t_k}) d\chi. \end{aligned} \tag{11}$$

Plugging (9) in (11), we get

$$\begin{aligned} F(w^*_{t_k}, t_{k+1}) &= r_\epsilon - F_w(\bar{w}_{t_k}, t_k)(\bar{w}_{t_{k+1}} - \bar{w}_{t_k}) \\ &\quad + \int_0^1 F_w\left(\bar{w}_{t_k} + \chi(w^*_{t_k} - \bar{w}_{t_k}), t_{k+1}\right)(w^*_{t_k} - \bar{w}_{t_k}) d\chi. \end{aligned} \tag{12}$$

From (12) and (10) and bounding we can obtain,

$$\begin{aligned} \|\bar{w}_{t_{k+1}} - w^*_{t_{k+1}}\| &\leq \kappa_\psi \kappa_\epsilon \\ &\quad + \kappa_\psi L_{F_w} L_w \Delta t \left(\kappa_r + \frac{1}{2} L_w \Delta t + \Delta t\right) \\ &\quad + \kappa_\psi L_{F_w} \kappa_r \left(\frac{1}{2}\kappa_r + \Delta t\right), \end{aligned}$$

where $\kappa_\psi = \frac{1}{\|F_w(\bar{w}_{t_k}, t_k)\|}$. For stability we require $\|\bar{w}_{t_{k+1}} - w^*_{t_{k+1}}\| \leq \kappa_r$. This implies

$$\begin{aligned} \kappa_r \geq & \kappa_\psi \kappa_\epsilon + \kappa_\psi L_{F_w} L_w \Delta t \left(\kappa_r + \frac{1}{2} L_w \Delta t + \Delta t\right) \\ & + \kappa_\psi L_{F_w} \kappa_r \left(\frac{1}{2}\kappa_r + \Delta t\right). \end{aligned}$$

Rearranging, we have

$$\begin{aligned} \left(1 - \frac{1}{2} L_{F_w} \kappa_\psi \kappa_r\right) \kappa_r \geq & \kappa_\psi \kappa_\epsilon + \kappa_\psi L_{F_w}(L_w + 1) \Delta t\, \kappa_r \\ & + L_{F_w} \kappa_\psi L_w \left(\frac{1}{2} L_w + 1\right) \Delta t^2. \end{aligned}$$

Stability follows if $\left(1 - \frac{1}{2} L_{F_w} \kappa_\psi \kappa_r\right) > 0$ and if there exist $\kappa \geq 0$ and Δt satisfying

$$\alpha_1^{NLE} \Delta t\, \kappa_r \leq \kappa \Delta t^2 \tag{13a}$$

$$\alpha_2^{NLE} \Delta t^2 + \kappa_\psi \kappa_\epsilon \leq \alpha_3^{NLE} \kappa_r, \tag{13b}$$

where $\alpha_1^{NLE}, \alpha_2^{NLE}$, and α_3^{NLE} are defined in the appendix of [14]. At every time t_k, $\bar{w}_{t_{k+1}}$ is obtained by solving (21). This is an approximation of $w^*_{t_{k+1}}$. The stability conditions guarantee that if $\kappa_r, \kappa_\epsilon = O(\Delta t^2)$, then the approximation error remains $O(\Delta t^2)$ for all $k > 0$. We have thus created an algorithm that tracks the solution manifold of $F(w,t) = 0$ stably by solving (within κ_ϵ) a single *truncated* Newton step per sampling time. This allows us to use *iterative linear algebra* schemes that can be terminated early.

In practice, one can still apply the above results to handle inequality constraints (bounds) by introducing *smoothing* schemes, as suggested in [10], [4], [13]. However, this leads to numerical instability. Note also that, in the presence introducing bounds, we can no longer invert algebraically the Newton system. In addition, non-smoothness prevents the direct application of standard calculus results. We resolve these technical difficulties in the following sections.

B. Linearized Generalized Equations

An important consequence of the structure of (7) is that it allows us to analyze the smooth and nonsmooth components independently. We start by defining the linearized generalized equation (LGE) at a given solution $w^*_{t_0}$,

$$r \in F(w^*_{t_0}, t_0) + F_w(w^*_{t_0}, t_0)(w - w^*_{t_0}) + \mathcal{N}_K(w). \quad (14)$$

If $K = \Re^n_+$, solving this LGE is equivalent to solving the perturbed linear complementarity problem,

$$w \geq 0,\ \nu = F(w^*_{t_0}, t_0) + F_w(w^*_{t_0}, t_0)\Delta w - r \geq 0,\ w^T\nu = 0. \quad (15)$$

If F_w is a symmetric matrix, then condition (15) are the optimality conditions of the QP problem,

$$\begin{aligned}\min_{\Delta w \geq -w^*_{t_0}} \quad & \frac{1}{2}\Delta w^T F_w(w^*_{t_0}, t_0)\Delta w \quad + F(w^*_{t_0}, t_0)^T\Delta w - r^T\Delta w.\end{aligned} \quad (16)$$

We can rewrite (7) at any point (w, t) in the neighborhood of $w^*_{t_0}$ in terms of (14) by defining the *residual*,

$$r(w, t) = F(w^*_{t_0}, t_0) + F_w(w^*_{t_0}, t_0)(w - w^*_{t_0}) - F(w, t). \quad (17)$$

This gives, for any point satisfying (7),

$$r(w, t) \in F(w^*_{t_0}, t_0) + F_w(w^*_{t_0}, t_0)(w - w^*_{t_0}) + \mathcal{N}_K(w). \quad (18)$$

This formulation will allow us to bound the distance between $(w^*_{t_0}, t_0)$ and neighboring points (w, t) in terms of $r(w, t)$.

Central to this study is the inverse operator $\psi^{-1} : Z \to W$ of the perturbed LGE (18), which we define as

$$\begin{aligned} & w \in \psi^{-1}[r] \\ & \Leftrightarrow\ r \in F(w^*_{t_0}, t_0) + F_w(w^*_{t_0}, t_0)(w - w^*_{t_0}) + \mathcal{N}_K(w). \end{aligned} \quad (19)$$

In other words, the operator is a multifunction from the space of the residual (perturbation) of the LGE to the space of the solution. Some basic properties arising from the definition of the inverse operator are as follows:

$$w^*_{t_0} \in \psi^{-1}[r(w^*_{t_0}, t_0)] = \psi^{-1}[0], \quad w^*_t \in \psi^{-1}[r(w^*_t, t)].$$

Definition 1: (*Strong Regularity* [12]). The GE (7) is said to be strongly regular at $w^*_{t_0}$ in the sense of Robinson if there exists a neighborhood $V_W \subseteq W$ of $w^*_{t_0}$ and a neighborhood $V_Z \subseteq Z$ of $r(w^*_{t_0}, t_0) = 0$, such that for every $r \in V_Z$, (18) has a unique solution $w = \psi^{-1}[r] \in V_W$, and the inverse mapping $\psi^{-1} : V_Z \to V_W$ is Lipschitz with constant L_ψ. That is, for any $r_1, r_2 \in V_Z$,

$$\|\psi^{-1}[r_1] - \psi^{-1}[r_2]\| \leq L_\psi \|r_1 - r_2\|.$$

This result is a generalization of the implicit function theorem for nonlinear equations. In other words, strong regularity guarantees the invertibility of the solution mapping. In Theorem 4.1 in [12] and Theorem 6 in [5]) it has been established that the so-called strong second-order conditions and the linear independence constraint qualification are sufficient to guarantee strong regularity of the NLP.

III. STABILITY OF APPROXIMATION ERROR

Using this basic set of tools, we now establish results that will allow us to construct algorithms for tracking the solution manifold of (7) approximately.

Theorem 1: (Theorem 2.3 in [12] and Theorem 3.3.4 in [6]) Assume (7) is strongly regular at $w^*_{t_0}$. Then, there exist neighborhoods V_W and V_T and a unique and Lipschitz continuous solution $w^*_t \in V_W$ of the GE (7) that satisfies, for each $t = t_0 + \Delta t \in V_T$,

$$\text{(i)} \quad \|w^*_t - w^*_{t_0}\| \leq L_w \Delta t \quad (20)$$

with $L_w > 0$. In addition, consider that $\bar{w}_t$ solves the truncated system

$$\delta_\epsilon \in F(w^*_{t_0}, t) + F_w(w^*_{t_0}, t_0)(\bar{w}_t - w^*_{t_0}) + \mathcal{N}_K(\bar{w}_t), \quad (21)$$

where r_ϵ is the solution residual satisfying $\|r_\epsilon\| \leq \delta_\epsilon > 0$. We have that $\bar{w}_t$ satisfies

$$\text{(ii)} \quad \|w^*_t - \bar{w}_t\| \leq L_\psi \left(\delta_\epsilon + \gamma(\Delta t)\Delta t\right),$$

with $\gamma(\Delta t) \to 0$ as $\Delta t \to 0$; and, if F_w is Lipschitz continuous, then

$$\text{(iii)} \quad \|w^*_t - \bar{w}_t\| \leq L_\psi \left(\delta_\epsilon + \kappa \Delta t^2\right)$$

with $\kappa > 0$.

Having a reference solution $w^*_{t_0}$, we can compute the approximate solution $\bar{w}_t$ by solving the LCP (15) or the QP (16) with $r = F(w^*_{t_0}, t_0) - F(w^*_{t_0}, t)$. From Theorem 1, we see that this approximation can be expected to be close to the optimal solution w^*_t. In our approximate algorithm, however, we relax the requirement that $w^*_{t_0}$ be available. Instead, we consider a linearization point $\bar{w}_{t_0}$ located in the neighborhood of $w^*_{t_0}$. In addition, we assume that the LCP is not solved exactly. In other words, $\bar{w}_t$ is the solution of the truncated system

$$r_\epsilon \in F(\bar{w}_{t_0}, t) + F_w(\bar{w}_{t_0}, t_0)(w - \bar{w}_{t_0}) + \mathcal{N}_K(w), \quad (22)$$

where $r_\epsilon \in \Re^n$ is the solution residual. This system can be posed in form (18) by using the following definition

$$\begin{aligned} r = r_\epsilon & + F(w^*_{t_0}, t_0) + F_w(w^*_{t_0}, t_0)(w - w^*_{t_0}) \\ & - F(\bar{w}_{t_0}, t) - F_w(\bar{w}_{t_0}, t_0)(w - \bar{w}_{t_0}). \end{aligned} \quad (23)$$

Note that, in this case, the perturbation r is an implicit function of the solution $w = \bar{w}_t$. In addition, we emphasize that (23) is used only as an analytical tool. In practice, however, (22) is solved.

Theorem 2: (Stability of Approximation Error). Assume (7) is strongly regular at $w^*_{t_0}$. Define $\bar{w}_t$ as the solution of the

perturbed LGE (22) where $\bar{w}_{t_0}$ is a point in the neighborhood V_W of $w^*_{t_0}$. The associated residual $r(\bar{w}_{t_0}, t_0)$ is assumed to satisfy

$$\|r(\bar{w}_{t_0}, t_0) - r(w^*_{t_0}, t_0)\| \leq \delta_r,$$

with $\delta_r > 0$. Assume there exists $\delta_\epsilon > 0$ such that $\|r_\epsilon\| \leq \delta_\epsilon$. If there exists $\kappa > 0$ and if Δt satisfies

$$\alpha_1^{GE} \Delta t\, \delta_r \leq \kappa \Delta t^2 \tag{24a}$$

$$(\alpha_2^{GE} + \kappa)\Delta t^2 + \delta_\epsilon \leq \alpha_3^{GE} \delta_r, \tag{24b}$$

with $\alpha_1^{GE}, \alpha_2^{GE}$ and α_3^{GE} defined in the Appendix of [14]; then the approximation error remains stable:

$$\|\bar{w}_{t_0} - w^*_{t_0}\| \leq L_\psi \delta_r \quad \Rightarrow \quad \|\bar{w}_t - w^*_t\| \leq L_\psi \delta_r.$$

Proof. See [14] □

Corollary 3: Assume conditions of Theorem 2 hold $\forall\, t_k \in [t_0, t_f]$. Then,

$$\|\bar{w}_{t_k} - w^*_{t_k}\| \leq L_\psi \delta_r,\ t_{k+1} = t_k + k \cdot \Delta t,\ \forall\, k \leq \frac{t_f - t_0}{\Delta t}.$$

Condition (24a) can be satisfied for $\delta_r = o(\Delta t), O(\Delta t^2)$. Condition (24b) is stricter. If $\delta_r = o(\Delta t)$, this condition states that the solution error should be at least $\delta_\epsilon = o(\Delta t)$. Note that a small L_ψ is beneficial because it relaxes both (24a) and (24b). The proposed scheme is equivalent to time-stepping methods used to solve differential variational inequalities (DVI) [11]. Finally, note that the above stability results can be applied directly to the NLP context since optimality conditions of QP (6) formulate an LGE of the form (22).

IV. AUGMENTED LAGRANGIAN STRATEGY

As we have seen, solving a single QP (6) at each time step is sufficient. However, it is crucial to have a fast solution strategy for the QP. Here, we propose to reformulate the NLP using an augmented Lagrangian (AL) function and solve the underlying QP using a projected successive over-relaxation (PSOR) strategy. To derive our strategy, we define the AL function,

$$\mathcal{L}_A(x, \bar{\lambda}, t, \rho) = f(x,t) + \bar{\lambda}^T c(x,t) + \frac{\rho}{2}\|c(x,t)\|^2. \tag{25}$$

A strategy to solve the original NLP (3) consists of computing solutions of the AL subproblem

$$\min_{x \geq 0} \quad \mathcal{L}_A(x, \bar{\lambda}, t, \rho) \tag{26}$$

for a sequence of increasing ρ. Note that the multipliers $\bar{\lambda}$ act as parameters of the AL subproblem. The solution of the subproblem is defined as $x^*(\bar{\lambda}, t)$. The multipliers can be updated externally as

$$\bar{\lambda} \leftarrow \bar{\lambda} + \rho\, c(x^*(\bar{\lambda}, t), t). \tag{27}$$

We thus define the solution pair $x^*(\bar{\lambda}, t)$, $\Lambda^*(\bar{\lambda}, t) = \bar{\lambda} + \rho\, c(x^*(\bar{\lambda}, t), t)$. The first-order conditions of (26) can be posed as a GE of the form

$$0 \in \nabla_x \mathcal{L}_A(x, \bar{\lambda}, t) + \mathcal{N}_{\Re^n_+}(x), \tag{28}$$

where

$$\nabla_x \mathcal{L}_A(x, \bar{\lambda}, t) = \nabla_x f(x,t) + (\bar{\lambda} + \rho\, c(x,t))^T \nabla_x c(x,t).$$

The linearized version of (28) defined at the NLP solution $x^*_{t_0}$, $\bar{\lambda} = \lambda^*_{t_0}$ is given by

$$\begin{aligned} r \in\ & \nabla_x \mathcal{L}_A(x^*_{t_0}, \lambda^*_{t_0}, t_0) \\ & + \nabla_{xx} \mathcal{L}_A(x^*_{t_0}, \lambda^*_{t_0}, t_0)(x - x^*_{t_0}) + \mathcal{N}_{\Re^n_+}(x) \end{aligned} \tag{29}$$

for $r = 0$. To establish perturbation results for the AL LGE in connection with those of the original NLP (3), we consider the following equivalent formulation of (28), proposed in [1]:

$$0 \quad \in \quad F(w, p(\bar{\lambda}), t) + \mathcal{N}_{\Re^n_+ \times \Re^m}(w), \tag{30}$$

where

$$F(w, p(\bar{\lambda}), t) = \begin{bmatrix} \nabla_x f(x,t) + \Lambda^T \nabla_x c(x,t) \\ c(x,t) + p(\bar{\lambda}) + \frac{1}{\rho}(\lambda^*_{t_0} - \Lambda) \end{bmatrix}, \tag{31}$$

$w^T = [x^T\, \Lambda^T]$, and

$$p(\bar{\lambda}) = \frac{1}{\rho}(\bar{\lambda} - \lambda^*_{t_0}). \tag{32}$$

For $t = t_0$ and $\bar{\lambda} = \lambda^*_{t_0}$, we have $p(\bar{\lambda}) = 0$, $x^*(p(\bar{\lambda}), t) = x^*_{t_0}$, and $\Lambda^*(p(\bar{\lambda}), t) = \lambda^*_{t_0}$. The solution of GE (30) is denoted as $w^*(p(\bar{\lambda}), t)$. The linearized version of (30) at $w^*_{t_0}$ is

$$r \in F(w^*_{t_0}, 0, t_0) + F_w(w^*_{t_0}, 0, t_0)(w - w^*_{t_0}) + \mathcal{N}_{\Re^n_+ \times \Re^m}(w), \tag{33}$$

where

$$F_w(w^*_{t_0}, 0, t_0) = \begin{bmatrix} \nabla_{xx} \mathcal{L}(w^*_{t_0}, t_0) & \nabla_x c(x^*_{t_0}, t_0) \\ \nabla_x^T c(x^*_{t_0}, t_0) & -\frac{1}{\rho}\mathbb{I}_m \end{bmatrix}. \tag{34}$$

We emphasize that the reformulation (30) is considered only for theoretical purposes. In practice, (28) is solved. We now establish the following approximation results in the context of the AL framework.

Lemma 4: Assume (28) is strongly regular at $w^*_{t_0}$. Then, there exist neighborhoods V_W,V_T, and V_p where the solution of the AL subproblem (26) satisfies, for each $t = t_0 + \Delta t \in V_T$, $p(\bar{\lambda}) \in V_p$,

$$(i) \quad \|w^*(\bar{\lambda}, t) - w^*_{t_0}\| \leq \frac{L_w}{\rho}\|\bar{\lambda} - \lambda^*_{t_0}\| + L_w \Delta t. \tag{35}$$

Furthermore, consider the approximate solution $\bar{x}(\bar{\lambda}, t)$ obtained from the perturbed LGE (29) with

$$r = \nabla_x \mathcal{L}_A(x^*_{t_0}, \lambda^*_{t_0}, t_0) - \nabla_x \mathcal{L}_A(x^*_{t_0}, \bar{\lambda}, t), \tag{36}$$

and associated multiplier $\bar{\Lambda}(\bar{\lambda}, t) = \bar{\lambda} + \rho\, c(\bar{x}(\bar{\lambda}, t), t)$. The pair, denoted by $\bar{w}(\bar{\lambda}, t)$, satisfies

$$(ii) \quad \|\bar{w}(\bar{\lambda}, t) - w^*(\bar{\lambda}, t)\| = O\left(\left(\Delta t + \frac{1}{\rho}\|\bar{\lambda} - \lambda^*_{t_0}\|\right)^2\right). \tag{37}$$

Proof. The result follows from the equivalence between (28) and (30), by recalling that $p(\lambda^*_{t_0}) = 0$, $p(\bar{\lambda}) = \frac{1}{\rho}\|\bar{\lambda} - \lambda^*_{t_0}\|$, and by applying Theorem 1. □

This result states that the solution of a perturbed AL LGE formed at $w^*_{t_0}$ provides a second-order approximation of the subproblem solution $w^*(\bar{\lambda}, t)$. The impact of the multiplier error can be made arbitrarily small by fixing ρ to a sufficiently large value.

Theorem 5: (Stability of Approximation Error for Augmented Lagrangian). Assume $w^*_{t_0}$ is a strongly regular solution of (29). Define $\bar{x}(\bar{\lambda}, t)$ as the solution of the LGE,

$$\begin{aligned} r_\epsilon \in & \nabla_x \mathcal{L}_A(\bar{x}_{t_0}, \bar{\lambda}, t) \\ & + \nabla_{xx} \mathcal{L}_A(\bar{x}_{t_0}, \bar{\lambda}, t_0)(x - \bar{x}_{t_0}) + \mathcal{N}_{\Re^n_+}(x), \end{aligned} \tag{38}$$

with an associated multiplier update $\bar{\Lambda}(\bar{\lambda}, t) = \bar{\lambda} + \rho\, c(\bar{x}(\bar{\lambda}, t), t)$. The pair is denoted by $\bar{w}(\bar{\lambda}, t)$. The reference linearization point $\bar{w}^T_{t_0} = [\bar{x}^T_{t_0}, \bar{\Lambda}^T_{t_0}]$ with $\bar{\Lambda}_{t_0} = \bar{\lambda} + \rho\, c(\bar{x}_{t_0}, t_0)$ is assumed to exist in the neighborhood V_W of $w^*_{t_0}$. The associated residual $r(\bar{w}_{t_0}, t_0)$ is assumed to satisfy $\|r(\bar{w}_{t_0}, t_0) - r(w^*_{t_0}, t_0)\| \leq \delta_r$ with $\delta_r > 0$. Furthermore, assume there exists $\delta_\epsilon > 0$ such that $\|r_\epsilon\| \leq \delta_\epsilon$. If there exists $\kappa > 0$, Δt and ρ satisfying

$$\alpha^{AL}_1 \Delta t \delta_r + \frac{L_w}{\rho}\left(\delta_r + \frac{L_w}{L_\psi}\Delta t\right) \leq \kappa \left(\Delta t + \frac{L_\psi \delta_r}{\rho}\right)^2 \tag{39a}$$

$$\alpha^{AL}_2 \left(\Delta t + \frac{L_\psi \delta_r}{\rho}\right)^2 + \delta_\epsilon \leq \alpha^{AL}_3 \delta_r, \tag{39b}$$

where $\alpha^{AL}_1, \alpha^{AL}_2, \alpha^{AL}_3$ are defined in the appendix in [14], then the approximation error remains stable:

$$\|\bar{w}_{t_0} - w^*_{t_0}\| \leq L_\psi \delta_r \quad \Rightarrow \quad \|\bar{w}(\bar{\lambda}, t) - w^*_t\| \leq L_\psi \delta_r.$$

Proof. See [14]. □

The recursive stability result of Corollary 3 also applies in this context. Note that if $\rho \to \infty$, conditions (39a)-(39b) reduce to (24a)-(24b). Therefore, similar order results to those of Theorem 2 can be expected for sufficiently large ρ. Note also that the initial multiplier error (bounded by δ_r) always appears divided by ρ. This indicates that relatively large initial multiplier errors can be tolerated by increasing ρ. Nevertheless, note that the second term on the left-hand side of (39a) remains $o(\Delta t)$ even if $\delta_r = O(\Delta t^2)$. In other words, this condition is more restrictive than (24a).

To solve the QP associated to the LGE (38), we follow a PSOR approach. The QP has the form,

$$\min_{z \geq \alpha} \quad \frac{1}{2} z^T \mathbf{M} z + \mathbf{b}^T z. \tag{40}$$

Any solution of this QP solves the LCP,

$$\mathbf{M}z + \mathbf{b} \geq 0, \quad z - \alpha \geq 0, \quad (z - \alpha)^T (\mathbf{M}z + \mathbf{b}) = 0. \tag{41}$$

Consider the following PSOR algorithm adapted from [7], [9]:

PSOR Algorithm

Given $z^0 \geq \alpha$, compute for $k = 0, 1, ..., n_{iter}$,

$$\begin{aligned} z^{k+1}_i &= (1 - \omega) z^k_i - \frac{\omega}{\mathbf{M}_{ii}} \left(\sum_{j<i} \mathbf{M}_{ij} z^{k+1}_j + \sum_{j>i} \mathbf{M}_{ij} z^k_j - \mathbf{b}_i \right) \\ z^{k+1}_i &= \max\left(z^{k+1}_i, \alpha_i\right), \quad i = 1, ..., n, \end{aligned} \tag{42}$$

where $\omega \in (0, 2)$ is the relaxation factor.

Theorem 6: (Theorem 2.1 in [9]). Let $\mathbf{M}$ be symmetric positive definite. Then, each accumulation point of the sequence $\{z^k\}$ generated by (42) converges to a solution of the LCP (15). The rate of convergence is R-linear.

It is known that, for the SOR method for linear systems, in order to reduce the error by a factor of $1/10$, SOR with non optimal parameter ω requires $O(n)$ iterations [7]. Here, $n = \dim(z)$. We can now establish our algorithm, which we refer to as AugLag.

AugLag Algorithm

Given $\bar{x}_{t_0}, \bar{\lambda}_{t_0}, \Delta t, \rho, n_{iter}$,

1) Evaluate $\nabla_x \mathcal{L}_A(\bar{x}_{t_k}, \bar{\lambda}_{t_k}, t_{k+1}, \rho)$ and $\nabla_{xx} \mathcal{L}_A(\bar{x}_{t_k}, \bar{\lambda}_{t_k}, t_k, \rho)$.
2) Compute step $\Delta \bar{x}_{t_{k+1}}$ by applying n_{iter} PSOR iterations to (40) with:
$$\begin{aligned} \mathbf{M} &= \nabla_{xx} \mathcal{L}_A(\bar{x}_{t_k}, \bar{\lambda}_{t_k}, t_k, \rho), \\ \mathbf{b} &= \nabla_x \mathcal{L}_A(\bar{x}_{t_k}, \bar{\lambda}_{t_k}, t_{k+1}, \rho) \end{aligned}$$
3) Update primal variables $\bar{x}_{t_{k+1}} = \bar{x}_{t_k} + \Delta \bar{x}_{t_{k+1}}$ and multipliers $\bar{\lambda}_{t_{k+1}} = \bar{\lambda}_{t_k} + \rho\, c(\bar{x}_{t_{k+1}}, t_{k+1})$.
4) Set $k \leftarrow k + 1$.

The proposed AugLag strategy is attractive because it performs linear algebra and active-set identification tasks simultaneously, it can exploit warm-start information, and it has a favorable computational complexity.

V. Numerical Example

To illustrate the developments, we consider the model predictive control of a nonlinear CSTR. The optimal control formulation is given by

$$\begin{aligned} \min_{u(\tau)} \int_t^{t+T} & \left(w_T (z_T - z^{sp}_T)^2 + w_C (z_C - z^{sp}_C)^2 \right. \\ & \left. + w_u (u - u^{sp})^2\right) d\tau \\ \text{s.t. } & \frac{dz_C}{d\tau} = \frac{z_C - 1}{\theta} + k_0 \cdot z_C \cdot \exp\left[\frac{-E_a}{z_T}\right], \; z_C(0) = \tilde{z}_C(t) \\ & \frac{dz_T}{d\tau} = \frac{z_T - z^f_T}{\theta} - k_0 \cdot z_C \cdot \exp\left[\frac{-E_a}{z_T}\right] \\ & \qquad + \alpha \cdot u \cdot (z_T - z^{cw}_T), \; z_T(0) = \tilde{z}_T(t) \\ & z^{min}_C \leq z_C \leq z^{max}_C, \quad z^{min}_T \leq z_T \leq z^{max}_T \\ & u^{min} \leq u \leq u^{max}. \end{aligned}$$

The system involves two states $z(\tau) = [z_C(\tau), z_T(\tau)]$ and one control $u(\tau)$. The set-points are denoted by the superscript sp. For implementation, the optimal control problem is converted into an NLP by applying an implicit Euler discretization scheme with grid size $\Delta\tau = 0.25$. The NLP is parametric in the initial conditions, which are implicit functions of t. The initial conditions are denoted by $\tilde{z}_T(t)$ and $\tilde{z}_C(t)$. To apply the AugLag algorithm, we set the AL penalty parameter $\rho = 100$ and fix the number of PSOR iterations to 25. We initialize the algorithm by perturbing an

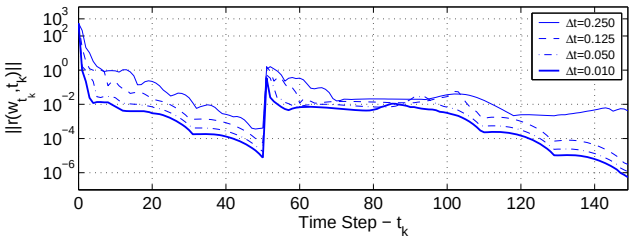

Fig. 1. Residual trajectories for AugLag with increasing Δt.

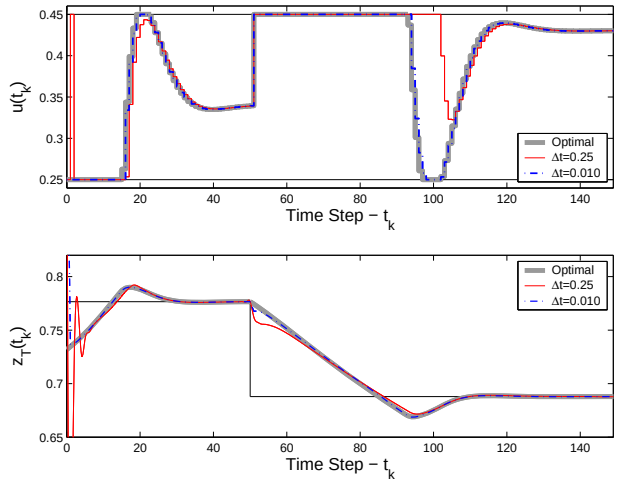

Fig. 2. AugLag and optimal trajectories for the control (top) and temperature (bottom).

initial solution $w^*_{t_0}$ as $\bar{w}_{t_0} \leftarrow w^*_{t_0} \cdot \delta_w$ where $\delta_w > 0$ is a perturbation. This generates the initial residual $r(\bar{w}_{t_0}, t_0)$. An additional perturbation, in the form of a set-point change, is introduced at $t_k = 50$. In Figure 1, we present the norm of the residuals along the simulation horizon with increasing Δt. Note that although the initial residual is large $O(10^3)$, the subsequent residuals remain stable and tend to decrease. The set-point change generates a residual that is only $O(10^0)$ and can be tolerated with no problem. The PSOR residuals r_ϵ at the beginning of the horizon are $O(10^{-1})$ and converge to $O(10^{-6})$ when the system reaches the set-points. In Figure 2, we present control and temperature profiles for $\Delta t = 0.25$ and $\Delta t = 0.01$. The approximation error decreases with the step size. The PSOR strategy identifies efficienctly active-set changes in subsequent steps. At a single step, up to 100 changes were observed. For the larger step size, note that even if the active-sets do not match, the residuals remain bounded and the system eventually converges to the optimal trajectories.

VI. CONCLUSIONS AND FUTURE WORK

We have presented new insights into enabling the implementation of NMPC at higher frequencies. The main idea is to solve a single, *truncated* quadratic programming problem per time step. We establish conditions guaranteeing that the approximation error remains stable even in the presence of active-set changes. In addition, we present truncated scheme that enables early termination and that performs linear algebra and active-set identification tasks simultaneously. As part of future work, we are exploring other algorithms with even cheaper steps. In addition, we seek to establish convergence results of approximate schemes.

ACKNOWLEDGMENTS

This work was supported by the U.S. Department of Energy through contract DE-AC02-06CH11357. We thank the reviewers for their questions and suggestions. These were valuable in improving the quality of the manuscript.

REFERENCES

[1] D. .Bertsekas. *Constrained Optimization and Lagrange Multiplier Methods*. Academic Press, 1982.
[2] F. Bonnans and A. Shapiro. *Perturbation Analysis of Optimization Problems*. Springer-Verlag, 2000.
[3] M. Diehl, H.G. Bock, and J.P. Schlöder. A real-time iteration scheme for nonlinear optimization in optimal feedback control. *SIAM J. Control and Optim.*, 43:1714–1736, 2005.
[4] M. Diehl, H. J. Ferreau, and N. Haverbeke. Efficient numerical methods for nonlinear MPC and moving horizon estimation. *In Nonlinear Model Predictive Control*, pages 391–417, 2009.
[5] A. L. Dontchev and T. R. Rockafellar. Characterizations of strong regularity for variational inequalities over polyhedral convex sets. *SIAM J. Optim.*, 6:1087–1105, 1996.
[6] J. Guddat, F. Guerra Vazquez, and H. T. Jongen. *Parametric Optimization: singularities, pathfollowing and jumps*. BG Teubner, 1990.
[7] M. Kočvara and J. Zowe. An iterative two-step algorithm for linear complementarity problems. *Numerische Mathematik*, 68:95–106, 1994.
[8] W. C. Li and L. T. Biegler. Process control strategies for constrained nonlinear systems. *Ind. Eng. Chem. Res.*, 27:1421–1433, 1988.
[9] O. L. Mangasarian. Solution of symmetric linear complementarity problems by iterative methods. *JOTA*, 22(4):465–485, 1977.
[10] T. Ohtsuka. A continuation/GMRES method for fast computation of non-linear receding horizon control. *Automatica*, 40:563–574, 2004.
[11] J. Pang and D. Stewart. Differential variational inequalities. *Mathematical Programming*, 113:345–424, 2008.
[12] S. M. Robinson. Strongly regular generalized equations. *Mathematics of Operations Research*, 5:43–61, 1980.
[13] A. G. Wills and W. P. Heath. Barrier function based model predictive control. *Automatica*, 40:1415–1422, 2004.
[14] V. M. Zavala and M. Anitescu. On-line nonlinear programming as a generalized equation. *Preprint ANL/MCS-P1641-0609*, Argonne National Laboratory, `http://www.mcs.anl.gov/uploads/cels/papers/P1641A.pdf`.